

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION  
EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété  
Intellectuelle  
Bureau international



(43) Date de la publication internationale  
13 décembre 2001 (13.12.2001)

PCT

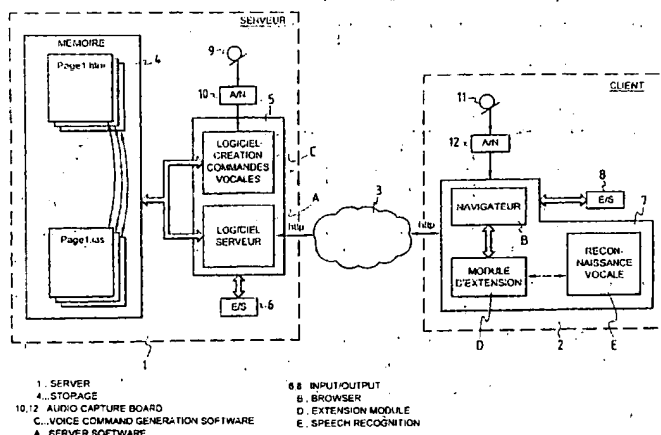
(10) Numéro de publication internationale  
**WO 01/95087 A1**

- (51) Classification internationale des brevets<sup>7</sup> : **G06F 3/16** (71) **Déposant (pour tous les États désignés sauf US) : INTERACTIVE SPEECH TECHNOLOGIES [FR/FR];**  
121, rue Chanzy, F-59260 Hellemmes (FR).
- (21) Numéro de la demande internationale : PCT/FR01/01560 (72) **Inventeurs; et**
- (22) Date de dépôt international : 21 mai 2001 (21.05.2001) (75) **Inventeurs/Déposants (pour US seulement) : GACHIE, Bruno [FR/FR]; 2, rue des Stations, F-59000 Lille (FR). DEWAVRIN, Anselme [FR/FR]; 38, Clairière du Mont, F-59250 Halluin (FR).**
- (25) Langue de dépôt : français
- (26) Langue de publication : français
- (30) Données relatives à la priorité : 00/07359 8 juin 2000 (08.06.2000) FR (74) **Mandataire : MATKOWSKA, Franck; Cabinet Beau de Loménie, 27 bis, rue du Vieux Faubourg, F-59800 Lille (FR).**

[Suite sur la page suivante]

(54) Title: VOICE-OPERATED SYSTEM FOR CONTROLLING A PAGE STORED ON A SERVER AND CAPABLE OF BEING DOWNLOADED FOR DISPLAY ON A CLIENT DEVICE

(54) Titre : SYSTEME DE COMMANDE VOCALE D'UNE PAGE STOCKEE SUR UN SERVEUR ET TELECHARGEABLE EN VUE DE SA VISUALISATION SUR UN DISPOSITIF CLIENT



(57) Abstract: The invention concerns a system for voice-operated control of a page designed to be displayed on a client device (2), which can exchange data with a remote server (1) via a telecommunication network (3), and which comprises means (11, 12) for recording a voice command spoken by a user, and speech recognition means enabling, from a recorded voice command, to determine and automatically control the execution of an action associated with said voice command. The server (1) comprises in storage, connected with said page (page (1).htm), at least a dictionary (page (1).ias) of one or several voice links comprising for each voice link at least an audio recording of the voice command; the client device is adapted to download in storage each dictionary associated with the page, and the speech recognition means of the client device (2) comprise a speech recognition programme (E) which is designed to carry out a comparison of the audio recording corresponding to the voice command with the audio recording(s) of each dictionary associated with the page.

(57) Abrégé : Le système permet la commande vocale d'une page destinée à être visualisée sur un dispositif client (2), qui d'une part peut échanger des données avec un serveur (1) distant via un réseau de télécommunication (3), et qui d'autre part comprend des moyens (11, 12) permettant l'enregistrement d'une commande vocale prononcée

[Suite sur la page suivante]

WO 01/95087 A1



(81) États désignés (*national*) : AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Publiée :**

— avec rapport de recherche internationale

(84) États désignés (*régional*) : brevet ARIPO (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), brevet eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen

En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.

par un utilisateur, et des moyens de reconnaissance vocale permettant, à partir d'une commande vocale enregistrée, de déterminer et de commander automatiquement l'exécution d'une action associée à cette commande vocale. Le serveur (1) comporte en mémoire, en lien avec ladite page (page (1).htm, ), au moins un dictionnaire (page (1).ias, ) d'un ou plusieurs liens vocaux comprenant pour chaque lien vocal au moins un enregistrement audio de la commande vocale; le dispositif client est apte à télécharger en mémoire chaque dictionnaire associé à la page, et les moyens de reconnaissance vocale du dispositif client (2) comportent un programme de reconnaissance vocale (E) qui est conçu pour réaliser une comparaison de l'enregistrement audio correspondant à la commande vocale avec le ou les enregistrements audio de chaque dictionnaire associé à la page.

**SYSTEME DE COMMANDE VOCALE D'UNE PAGE STOCKEE SUR UN  
SERVEUR ET TELECHARGEABLE EN VUE DE SA VISUALISATION  
SUR UN DISPOSITIF CLIENT**

5 La présente invention concerne la commande vocale de pages accessibles sur un serveur via un réseau de télécommunication, et plus particulièrement de pages hypertextes. Elle trouve principalement, mais non exclusivement, son application à la navigation hypertexte par commande vocale sur un réseau de télécommunication de type Internet.

10 Dans le présent texte, le terme « serveur » désigne d'une manière générale tout système informatique dans lequel sont stockées des données et qui est consultable à distance via un réseau de télécommunication.

Le terme « page » désigne tout document prévu pour être affiché  
15 sur un écran et stocké sur un site serveur à une adresse donnée.

Le terme « dispositif client » désigne d'une manière générale tout dispositif informatique apte à envoyer des requêtes à un site serveur pour que ce dernier lui envoie en retour les données objets de la requête, et en particulier un page déterminée, par exemple identifiée dans la requête  
20 par son adresse sur le serveur.

Le terme réseau de télécommunication désigne d'une manière générale tout moyen de communication permettant l'échange de données à distance entre un site serveur et un dispositif client ; il peut s'agir d'un réseau local (LAN) tel que l'intranet d'une entreprise ou encore d'un  
25 réseau longue distance (WAN) tel que par exemple le réseau Internet, ou encore d'un ensemble de réseaux de type différents et interconnectés.

Pour simplifier l'envoi à distance de pages entre un serveur et un dispositif client connecté à ce serveur via un réseau de télécommunication, on utilise couramment des systèmes de navigation  
30 hypertexte, qui permettent de naviguer parmi un ensemble de pages reliées les unes aux autres par des liens encore appelés liens hypertextes ou hyper-liens. En pratique, dans un système de navigation hypertexte,

une page hypertexte contient en plus du texte de base devant être affiché à l'écran, des caractères ou séquences de caractères particuliers qui peuvent faire ou non partie intégrante du texte de base, et qui constituent les liens hypertextes de la page. Lorsque que ces liens hypertextes font  
5 partie intégrante du texte de base de la page, ils sont différenciés des autres caractères de la page de base, par exemple en étant soulignés et/ou visualisés dans une autre couleur, etc. Pour la gestion de la navigation hypertexte, le dispositif client est de manière usuelle équipé d'un logiciel de navigation, encore appelé navigateur. Lors de la sélection  
10 par un utilisateur d'un lien hypertexte dans la page en cours de visualisation, le logiciel de navigation dans un premier temps établit et envoie automatiquement une requête au serveur, permettant à ce dernier d'envoyer la page attachée au lien hypertexte qui a été sélectionné, et dans un second temps affiche à l'écran la nouvelle page qui lui est  
15 envoyée par le serveur.

Afin de faciliter l'activation des liens hypertextes dans un système de navigation hypertexte, on a déjà proposé des système d'activation par commande vocale, dans lesquels le lien hypertexte est prononcé par l'utilisateur, et est reconnu automatiquement par un système de  
20 reconnaissance vocale. Ces systèmes d'activation vocale remplacent avantageusement les systèmes d'activation manuelle traditionnels (clavier/souris), et deviennent même indispensables dans toutes les applications où il n'est pas envisageable ou souhaité de mettre œuvre un outil manuel tel qu'un clavier ou une souris. Un exemple de ce type  
25 d'application est la navigation vocale sur le réseau mondial internet au moyen de téléphones portables WAP.

A ce jour, tous les systèmes d'activation vocale de liens dans une page hypertexte sont essentiellement basés sur une analyse automatique (« parsing ») de la page hypertexte, sur une détection automatique les  
30 liens présents dans la page, et sur la génération automatique de phonèmes à partir de chaque lien détecté.

Plus particulièrement dans le brevet US-A- 6,029,135 , on décrit un système de navigation hypertexte par commande vocale qui peut être mis en œuvre selon deux variantes : une première variante dite « run time » et une seconde variante dite « off line ». Dans la variante « off line », il est

5 enseigné de faire générer par le producteur de pages hypertextes, « *des données supplémentaires* » pour la commande vocale de ces pages, lesquelles données supplémentaires sont téléchargées avec la page hypertexte depuis le serveur. Ces « *données supplémentaires* » sont utilisées par le « client » pour effectuer la reconnaissance vocale des

10 mots prononcés par un utilisateur via un microphone, l'intelligence de reconnaissance vocale étant localisée au niveau du client. Dans l'unique mode de réalisation décrit, les « *données supplémentaires* » sont constituées par un dictionnaire de phonèmes, associé à un modèle de probabilité. Le dictionnaire de phonèmes et le modèle de probabilité

15 associé sont générés automatiquement à partir de la page par analyse automatique du contenu du document et extraction automatique des liens présents dans le document. A cet effet, un logiciel spécifique appelé « *manager* » est utilisé.

Les solutions de l'art antérieur et en particulier celle retenue dans

20 le brevet US-A- 6,029,135 présentent l'inconvénient majeur d'être basées sur une reconnaissance phonétique, ce qui d'une part complique la reconnaissance vocale, et est une source importante d'erreur , et ce qui d'autre part impose la mise en œuvre d'un logiciel complexe (« *manager* ») permettant la traduction automatique de chaque mot sous

25 forme de phonèmes, et l'élaboration automatique de modèle de probabilité pour la mise en œuvre de la reconnaissance phonétique. Le logiciel de traduction phonétique est d'autant plus complexe si l'on souhaite par exemple intégrer différentes prononciations d'un mot, pour tenir compte de la langue. Egalement, ce type de solution présente

30 l'inconvénient d'être dépendant d'une langue pour la transcription automatique du texte de la commande vers sa traduction en phonétique.

Pour les raisons ci-dessus, ces solutions sont à ce jour relativement onéreuses, sont réservées à des systèmes de navigation professionnels très spécialisés, et sont donc peu adaptées à des applications dites grand public.

- 5 La présente invention a pour principal objet un système qui permet la commande vocale d'une page qui est destinée à être visualisée sur un dispositif client apte à échanger des données avec un serveur distant via un réseau de télécommunication, et qui pallie les inconvénients précités des systèmes existants. Par commande vocale d'une page, on vise non  
10 seulement l'activation vocale de liens associés à la page, mais également et plus généralement l'activation par la voix de toute commande associée à la page visualisée, la commande n'étant pas nécessairement matérialisée par un mot visualisé sur l'écran du dispositif client mais pouvant être cachée. L'exécution de la commande associée à une page  
15 peut être de nature variée et n'est pas limitative de l'invention ( activation d'un lien hypertexte renvoyant sur une nouvelle page du serveur, commande de périphériques du dispositif client tel que par exemple une imprimante, ouverture ou fermeture de fenêtres sur le dispositif client, déconnexion du dispositif client, connexion du dispositif client sur un  
20 nouveau serveur etc...).

De manière connue, notamment par le brevet US-A- 6,029,135, le dispositif client comprend des moyens tels qu'un microphone et un carte d'acquisition audio, permettant l'enregistrement d'une commande vocale prononcée par un utilisateur, et des moyens de reconnaissance vocale  
25 permettant, à partir d'une commande vocale enregistrée, de déterminer et de commander automatiquement l'exécution une action associée à cette commande.

De manière caractéristique et essentielle selon l'invention, le serveur comporte en mémoire en lien avec ladite page au moins un  
30 dictionnaire d'un ou plusieurs liens vocaux comprenant pour chaque lien vocal au moins un enregistrement audio de la commande vocale ; le

dispositif client est apte à télécharger en mémoire chaque dictionnaire associé à la page, et les moyens de reconnaissance vocale du dispositif client comportent un programme de reconnaissance vocale qui est conçu pour réaliser une comparaison de l'enregistrement audio correspondant à

5 la commande vocale avec le ou les enregistrements audio de chaque dictionnaire associé à la page.

D'autres caractéristiques et avantages de l'invention apparaîtront plus clairement à la lumière de la description ci-après d'un exemple particulier de réalisation, laquelle description est donnée à titre

10 d'exemple non limitatif et en référence aux dessins annexés sur lesquels :

- la figure 1 est une représentation schématique des principaux éléments composant un système de commande vocale conforme à l'invention,

- la figure 2 représente les principales étapes d'un programme

15 d'aide à la création d'un dictionnaire de liens vocaux caractéristiques de l'invention, et de mise en relation du dictionnaire créé avec une page d'un serveur, en vue de la commande vocale de cette page,

- les figures 3 à 6 sont des exemples de fenêtres générées par le programme d'aide à la création de dictionnaires,

20 - la figure 7 illustre les principales étapes mises en œuvre par un dispositif client dans la phase de téléchargement d'un dictionnaire associé à une page fournie par un serveur,

- la figure 8 illustre les principales étapes mises en œuvre par le programme de reconnaissance vocale exécuté en local par le dispositif

25 client.

En référence à la figure 1, dans un exemple particulier de réalisation, l'invention met en œuvre un serveur informatique 1, sur lequel peuvent se connecter un ou plusieurs dispositifs clients 2 via un réseau de télécommunication 3. Plus particulièrement, dans l'exemple illustré, le

30 serveur informatique 1 héberge de manière usuelle un ou plusieurs sites web, et les dispositifs clients sont conçus pour se connecter sur le serveur

1 via le réseau mondial Internet, et pour échanger des données avec ce serveur selon le protocole de communication IP usuel.

Chaque site web hébergé par le serveur 1 est constitué d'une pluralité de pages html matérialisées par des fichiers au format .htm  
5 (figure 1 /'page1.htm, etc...) et reliés entre elles par des hyper-liens. Ces pages sont stockées de manière usuelle dans une unité mémoire 4 accessible en lecture et écriture par l'unité de traitement 5 du serveur 1. Outre l'unité mémoire 4 et l'unité de traitement 5, le serveur 1 comporte également de manière usuelle des moyens d'entrée/sortie 6, comprenant  
10 au moins un clavier permettant à un administrateur du serveur de saisir des données et/ou commandes, et au moins un écran permettant la visualisation des données du serveur, et notamment la visualisation des pages d'un site. Pour la gestion de l'échange de données avec un client 2 via le réseau 3, l'unité de traitement 5 comporte en mémoire vive un  
15 logiciel serveur A connu en soit, et permettant notamment l'envoi à un client 2 connecté au serveur 1 du ou des fichiers correspondant à la requête du client.

Un dispositif client 2 comporte de manière connue une unité de traitement 7 apte à être connectée via une interface de communication  
20 au réseau 3, et connectée également à des moyens d'entrée/sorties 8, dont au moins un écran pour la visualisation de chaque page html envoyée par le serveur 1. L'unité de traitement met en œuvre un logiciel de navigation B connu en soit, encore appelé navigateur (par exemple le logiciel de navigation de Netscape).

25 L'invention, dont les moyens nouveaux vont à présent être détaillés à la lumière d'un exemple particulier de réalisation, n'est pas limitée à une application de type Internet, mas peut d'une manière plus générale être appliquée à toute architecture client/serveur indépendamment du type de réseau de télécommunication et du protocole d'échange de données  
30 utilisés. Egalement, le dispositif client peut indifféremment être un poste fixe, ou comprendre une unité mobile tel qu'un téléphone portable, de



type WAP, permettant un accès au réseau de télécommunication 3.

L'invention repose de manière essentielle sur la mise en œuvre, pour chaque page du serveur à laquelle on souhaite associer une fonction de commande vocale, d'au moins un dictionnaire de liens  
5 vocaux, qui est stocké en mémoire du serveur 1 en lien avec ladite page, et qui présente la particularité de contenir pour chaque commande vocale, au moins un enregistrement audio, de préférence sous forme compressée, de la commande vocale. Dans l'exemple illustré sur la figure 1, à chaque page html est associé en mémoire un unique dictionnaire  
10 matérialisé par un fichier comportant le même nom que la page mais avec une extension différente arbitrairement désignée « .ias » dans la suite de la présente description. Ainsi, à la page html matérialisée par le fichier page 1.htm est associé, en mémoire du serveur 1, le fichier dictionnaire page1.ias, etc... Dans une autre variante, il est envisageable d'associer  
15 plusieurs dictionnaires à une même page.

Pour permettre la construction des fichiers dictionnaires ( .ias), le serveur 1 est équipé d'un microphone 9 connecté à une carte d'acquisition audio 10 (connue en soit), qui d'une manière générale permet de transformer le signal analogique délivré par le microphone 9 en une  
20 information de type numérique. Cette carte d'acquisition audio 10 communique avec l'unité de traitement 5 du serveur 1, et permet à ce dernier d'effectuer l'acquisition via le microphone 9 d'enregistrements vocaux sous forme numérique. L'unité de traitement 5 est en outre apte à exécuter un logiciel C spécifique de l'invention, dont une variante sera  
25 décrite ci-après, et qui permet d'assister le créateur d'un site web dans la construction des dictionnaires de liens vocaux.

De manière similaire, afin de permettre l'acquisition par l'unité de traitement 7 d'un dispositif client 2 d'une commande vocale prononcée par l'utilisateur, ledit dispositif client 2 est également équipé d'un  
30 microphone 11 et d'une carte d'acquisition audio 12. Tel que cela sera expliqué de manière plus détaillée ultérieurement, la reconnaissance

vocale automatique d'une commande vocale prononcée par l'utilisateur du dispositif client 2, en liaison avec une page en cours de visualisation sur l'écran du dispositif client 2, est réalisée en local par l'unité de traitement 7 du dispositif client 2, après téléchargement du fichier dictionnaire associé à la page en cours de visualisation.

### Spécifications d'un Fichiers Dictionnaire (.ias)

Dans un exemple de réalisation, un fichier dictionnaire contient un ou plusieurs liens vocaux enregistrés les uns à la suite des autres, chaque lien vocal possédant plusieurs attributs concaténés :

1. le nom (qui correspond au mot phonétique de la commande vocale qui doit être prononcé par l'utilisateur pour activer le lien) ;
2. le type
3. l'adresse (plus communément appelée URL) permettant de localiser sur le serveur la ressource associée à la commande vocale ;
4. la cible ( c'est-à-dire le nom de la fenêtre dans la quelle doit d'afficher la nouvelle page) ;
5. un enregistrement audio ( encore appelé modèle acoustique) masculin
6. un enregistrement audio ( encore appelé modèle acoustique) féminin

L'attribut « type » d'un lien vocal est utilisé notamment pour spécifier :

- qu'il s'agit bien d'un lien vocal et pour le différencier par exemple des hyper lien d'une page html sans possibilité de commande vocale,
- s'il s'agit d'un lien dont le nom apparaît dans le texte de la page associée ,
- si ce lien doit être caché ou si au contraire le nom du lien peut être affiché sur l'écran du dispositif client 2 dans une fenêtre spécifique contenant pour l'utilisateur les noms de tous les liens (non cachés) qu'il peut activer par la voix.

Plus particulièrement, à titre d'exemple, en langage C++, un lien vocal peut être transcrit de la manière suivante :

| Information                          | type C | Taille en octets   | Taille maximale | Valeurs permises |
|--------------------------------------|--------|--------------------|-----------------|------------------|
| Type de lien                         | DWORD  | 4                  | 4               | Voir plus bas    |
| Taille du nom                        | short  | 2                  | 2               | nombre positif   |
| Nom                                  | chars  | taille du nom      | 200             | caractères ANSI  |
| Taille du lien URL                   | short  | 2                  | 2               | nombre positif   |
| URL                                  | chars  | taille du lien URL | 2048            | caractères ANSI  |
| Taille de la cible                   | short  | 2                  | 2               | nombre positif   |
| Cible                                | chars  | taille de la cible | 200             | caractères ANSI  |
| Taille du modèle acoustique masculin | short  | 2                  | 2               | nombre positif   |
| Modèle acoustique masculin           | chars  | taille du modèle   | 2048            | toutes           |
| Taille du modèle acoustique féminin  | short  | 2                  | 2               | nombre positif   |
| Modèle acoustique féminin            | chars  | taille du modèle   | 2048            | toutes           |

5

#### Programme de construction d'un fichier Dictionnaire (figure 2)

- 10 Les principales étapes du programme de création d'un fichier dictionnaire vont à présent être expliquées en référence principalement à la figure 2. Dans l'exemple de la figure 1, ce programme est exécuté par l'unité de traitement 5 du serveur, après que l'administrateur du serveur ait choisi l'option correspondante permettant le lancement du programme.
- 15 Néanmoins, dans une autre application, ce programme pourra

avantageusement être mis à la disposition du créateur d'un site web, en étant mis en œuvre sur une machine différente du serveur, les fichiers dictionnaires (.ias) créés à l'aide de ce programme ainsi que les pages du sites web étant ensuite téléchargés dans l'unité mémoire 4 du serveur 2.

5        En référence à la figure 2, la création d'un fichier dictionnaire page (m).ias associé à une page html commence (étape 201) par l'ouverture du fichier page (m).htm de la page, suivie d'une extraction automatique des hyper liens présents dans la page (étape 202) et de la création d'un fichier dictionnaire page(m).ias avec ouverture d'une fenêtre de  
10 visualisation et de modification et/ou saisie de liens vocaux de ce dictionnaire (fenêtre « Dictionnaire » / étape 203). On a représenté sur la figure 3, un exemple de fenêtre créée à l'issue de l'étape 203. Dans cet exemple, ont été détectés et extraits automatiquement de la page(m).htm trois hyper liens, et pour chacun des ces hyper-liens a été créé  
15 automatiquement dans le dictionnaire associé page(m).ias, un lien vocal dont l'attribut adresse contient l'adresse URL de l'hyper lien correspondant retrouvé automatiquement dans le fichier page (m).htm.

A partir de cette première fenêtre (figure 3), il est possible soit de sélectionner dans la fenêtre de la figure 3 un lien existant dans le  
20 dictionnaire ( étape 204), soit de créer un nouveau lien vocal dans le dictionnaire (étape 205) en sélectionnant la commande appropriée dans un menu géré par la fenêtre de la figure 3.

Il convient ici de souligner que la fonction de création d'un nouveau lien vocal permet avantageusement de créer une commande vocale, qui  
25 ne correspond pas nécessairement à un hyper lien présent dans la page, et par là-même offre la possibilité de programmer des commandes vocales variées, et qui plus est des commandes cachées. Egalement, l'étape précitée d'extraction automatique (étape 202) est facultative, et ne se justifie que par un soucis de faciliter et d'accélérer la création du  
30 dictionnaire, en évitant à l'utilisateur de devoir créer manuellement dans le dictionnaire les liens vocaux correspondant à des hyper-liens de la

page et de devoir saisir les adresses URL correspondantes.

En cas de sélection d'un lien vocal existant ou de création d'un nouveau lien vocal, le programme ouvre une deuxième fenêtre « propriétés du lien » du type de celle illustrée sur la figure 4 (étape 206),  
5 qui permet à l'utilisateur de saisir et/ou modifier les attributs précédemment décrit d'un lien vocal.

En particulier, dans cette fenêtre l'utilisateur a la possibilité de sélectionner un premier bouton d'action « Enreg » pour l'enregistrement d'une commande vocale par une voix masculine et un second bouton  
10 d'action « Enreg » pour l'enregistrement d'une commande vocale par une voix féminine. Lorsque l'utilisateur sélectionne l'un des boutons d'action précités, le programme exécute automatiquement un module d'acquisition d'un enregistrement audio. Ce module, une fois lancé permet l'acquisition via le microphone 9 d'un enregistrement audio sous forme numérique de  
15 la commande vocale ( voix masculine ou féminine selon le cas) pendant un laps de temps donné contrôlé, et à l'issue de ce laps de temps réalise une compression automatique de cet enregistrement par tout procédé connu de compression de données, puis sauvegarde cet enregistrement audio compressé dans le fichier dictionnaire page(m).ias

20 Une fois que l'utilisateur a validé que toutes les propriétés d'un lien vocal ont été saisies ou modifiées, le programme ferme la fenêtre « propriétés du lien » correspondante (étape 207), et une fois que tous les liens vocaux du dictionnaire page (m).ias ont été créés de manière complète, l'utilisateur commande la fermeture de la fenêtre  
25 « Dictionnaire », et par là-même la fermeture du dictionnaire page (m).ias (étape 208). La figure 5 illustre un exemple de fenêtre « propriété du lien » pour le lien vocal « Supérieur » mise à jour avant fermeture de la fenêtre ; la figure 6 illustre un exemple de fenêtre « Dictionnaire » mise à jour avant fermeture du dictionnaire page (m).ias.

30 Une fois un dictionnaire page(m).ias complètement créé, le programme crée automatiquement (étape 209) un lien entre la page

(fichier page(m).htm ) et le dictionnaire associé (fichier page(m).ias). et ferme le fichier dictionnaire (page(m).ias). Dans une variante de réalisation, ce lien est créé par insertion du nom (page(m).ias) du dictionnaire associé dans le fichier (page(m).htm) de la page. Un exemple

5 d'implémentation du fichier page(m).htm est donné ci-dessous :

```
<html>
```

```
<head>
```

```
<TITLE> ( titre du fichier de la page html) </TITLE>
```

```
</head>
```

10

```
<body>
```

```
<a href = <"suivante.htm"> Suivante</a><br>
```

```
<a href = <'precedente.htm"> Précédente</a><br>
```

15 <a href = <"superieur.htm"> Supérieur</a><br>

```
...
```

```
<p><embed src="page(m).ias" pluginspage="" type="application/x-
NavigationByVoice" width="120" height="50"></embed></p>
```

20 </body>

```
</html>
```

La phase de transmission d'un dictionnaire entre le serveur 1 et un  
25 dispositif client 2 ainsi que la phase de reconnaissance vocale vont à  
présent être détaillées en référence aux figures 1, 7 et 8.

#### Transmission d'un dictionnaires (.ias)

30 Initialement à l'aide du programme navigateur (B), le dispositif client 2  
demande au serveur 1 de lui envoyer une page html ( par exemple le

fichier page(m).htm) De manière usuelle, le navigateur (B) analyse le fichier page(m).htm et affiche sur l'écran le contenu de la page au fur et à mesure où il reçoit les données afférentes à cette page ( figure 7/ étape 701).

- 5            Au cours de l'analyse automatique du fichier page(m).htm, lorsque le navigateur détecte l'information selon laquelle un dictionnaire est attaché à cette page ( détection de src="page(m).ias" dans le fichier), il charge un module d'extension D (figure 1) stocké en mémoire vive du dispositif client (étape 702), et parallèlement lance un programme de reconnaissance vocale également stocké en mémoire vive, pour le cas où ce programme n'aurait pas déjà été lancé ( ce qui est le cas par exemple la première fois où au cours d'une session une page (.htm) avec dictionnaire (.ias) attaché est reçue par le dispositif client 2).

- 15           Le navigateur envoie ensuite une requête au serveur 1 (étape 703) afin que celui ci lui transmette le fichier dictionnaire page(m).ias identifié dans le fichier page(m).htm.

Après réception par le dispositif client 2 du fichier dictionnaire page(m).ias, le navigateur (B) du dispositif client 2 envoie le fichier dictionnaire au module d'extension (D) (étape 705).

- 20           Ce module d'extension (D) crée à son tour un lien entre le fichier dictionnaire page(m).ias et le programme de reconnaissance vocale (E) (étape 706). Ensuite (étape 707), le module d'extension (D) analyse le contenu du fichier dictionnaire page(m).ias et affiche à l'écran pour l'utilisateur, par exemple dans une nouvelle fenêtre, les noms (attribut « nom ») de tous les liens vocaux du fichier dictionnaire page(m).ias, pour lesquels la valeur de l'attribut « type » autorise une visualisation (commandes vocales non cachées). (étape 706).
- 25

### Reconnaissance vocale

Cette fonction est assurée par le programme de reconnaissance vocale (E), à partir d'une commande vocale entrée par l'utilisateur au moyen du microphone 11 et par comparaison avec le ou les fichiers dictionnaires avec lequel un lien a été établi. Il convient ici de souligner que le programme de reconnaissance vocale peut être lancé avec simultanément plusieurs module d'extension actifs.

Plus particulièrement, en référence à la figure 8, une fois lancé le programme de reconnaissance vocale (E) est dans l'attente de la détection d'un son par le microphone 11. Lorsque l'utilisateur du dispositif client prononce une commande à la voix, celle-ci est enregistrée automatiquement sous forme numérique ( étape 801), et le programme de reconnaissance vocal effectue une compression de cette enregistrement, en appliquant la même méthode de compression que celle utilisée par le programme (C) de création de dictionnaires. Ensuite (étape 803), le programme de reconnaissance vocale (E) effectue automatiquement une comparaison des données numériques correspondant à cet enregistrement audio compressé avec les données numériques de chaque enregistrement audio compressé (modèles acoustiques masculins et féminins) du fichier dictionnaire page (m).ias. (ou d'une manière générale de tous les fichiers dictionnaires pour lesquels un lien avec le programme de reconnaissance vocale est actif), en vue d'en déduire automatiquement le lien vocal du dictionnaire correspondant à la commande prononcée par l'utilisateur.

Plus particulièrement, dans une variante de réalisation de l'invention chaque comparaison des enregistrements audio compressés est mise en œuvre selon la méthode DTW (Dynamic Time Warping), et donne pour résultat une note de reconnaissance caractérisant le similitude entre les enregistrements. Seule la note la plus élevée est ensuite retenue par le programme de reconnaissance vocale, et



comparée avec un seuil de détection prédéterminé en dessous duquel il est considéré que le mot prononcé n'a pas été reconnu comme une commande vocale. Si la note la plus élevée résultant des comparaisons précitées est supérieure à ce seuil, le programme de reconnaissance  
5 vocale reconnaît automatiquement le lien vocal correspondant à cette note comme étant la commande vocale prononcée par l'utilisateur.

De manière avantageuse selon l'invention, la reconnaissance vocale étant basée sur une comparaison d'enregistrements audio numériques (enregistrements audios des liens vocaux d'un dictionnaire  
10 .ias et de l'enregistrement audio de la commande vocale prononcée par l'utilisateur), on simplifie et on fiabilise très largement la reconnaissance vocale, comparativement à des systèmes de reconnaissance de type phonétique tel que celui mis en œuvre dans le brevet US-A- 6,029,135. Egalement, on s'affranchit de toute dépendance à une langue particulière.

15 Après reconnaissance d'un lien vocal, le programme de reconnaissance vocale envoie au navigateur (B) (étape 804) l'action qui est associée à ce lien vocal et qui est codée dans le dictionnaire, c'est-à-dire dans l'exemple particulier décrit précédemment l'adresse URL de ce lien vocal.

20 Si l'action associée correspond au chargement et à l'affichage d'une nouvelle page identifiée par son adresse URL, le navigateur (B), avant l'envoi de la requête appropriée au serveur, décharge la page en cours de visualisation ( page(m).htm ) ainsi que le module d'extension qui lui est associé; lequel module d'extension avant téléchargement casse le  
25 lien établi entre le programme de reconnaissance vocale (E) et le fichier dictionnaire page(m).ias. Ensuite les étapes de fonctionnement sont reprises à l'étape (701) précitée.

Dans l'exemple particulier de réalisation qui a été décrit, chaque lien vocal est caractérisé par une adresse (URL), qui est communiquée au  
30 navigateur du dispositif client lorsque ce lien vocal a été reconnu par le programme de reconnaissance vocal, ce qui permet ensuite au navigateur

de dialoguer avec le serveur afin que ce dernier envoie au dispositif client la ressource correspondant à cette adresse, et par exemple une nouvelle page. Ceci n'est toutefois pas limitatif de l'invention. On peut généraliser l'utilisation de cet attribut « adresse » d'un lien vocal pour coder d'une manière générale l'action qui est associée à la commande vocale définie par le lien vocal, et qui doit être automatiquement exécutée lors de la reconnaissance automatique d'un lien vocal par le programme de reconnaissance vocale. Ainsi, cette action codée dans l'attribut « adresse », peut non seulement être une adresse localisant une ressource stockée sur le serveur 1, mais pourrait également être une adresse localisant une ressource (données, programme exécutable, ...) stockée en local au niveau du dispositif client 2, ou un code de commande d'une action exécutable par le dispositif client tel que par exemple, et de manière non limitative, la commande d'un périphérique en local au niveau du dispositif client (impression d'un document, ouverture ou fermeture d'une fenêtre sur l'écran du dispositif client, arrêt de la communication avec le serveur et éventuellement mise en communication avec un nouveau serveur dont l'adresse serait spécifiée dans l'attribut « adresse », déconnexion définitive du dispositif client par rapport au réseau de télécommunication 3, etc...).

## REVENDEICATIONS

1. Système pour la commande vocale d'une page destinée à être  
5 visualisée sur un dispositif client (2), qui d'une part peut échanger des données avec un serveur (1) distant via un réseau de télécommunication (3), et qui d'autre part comprend des moyens (11,12) permettant l'enregistrement d'une commande vocale prononcée par un utilisateur, et des moyens de reconnaissance vocale permettant,  
10 à partir d'une commande vocale enregistrée, de déterminer et de commander automatiquement l'exécution d'une action associée à cette commande vocale, caractérisé en ce que le serveur (1) comporte en mémoire, en lien avec ladite page, au moins un dictionnaire d'un ou plusieurs liens vocaux comprenant pour chaque lien vocal au moins un  
15 enregistrement audio de la commande vocale, en ce que le dispositif client est apte à télécharger en mémoire chaque dictionnaire associé à la page, et en ce que les moyens de reconnaissance vocale du dispositif client (2) comportent un programme de reconnaissance vocale (E) qui est conçu pour réaliser une comparaison de  
20 l'enregistrement audio correspondant à la commande vocale avec le ou les enregistrements audio de chaque dictionnaire associé à la page.
2. Système selon la revendication 1 caractérisé en ce qu'un lien vocal comporte plusieurs enregistrements audio de la commande vocale, dont au moins un enregistrement d'une voix féminine et un  
25 enregistrement d'une voix masculine.
3. Système selon la revendication 1 ou 2 caractérisé en ce que la page du serveur (1) comporte une information identifiant le ou les dictionnaires associés, et en ce que le dispositif client (2) est conçu d'une part pour détecter cette information lors de la visualisation de la  
30 page, et d'autre part en cas de détection de cette information pour envoyer une requête au serveur (1) afin que ce dernier lui envoie le

dictionnaire identifié par cette information.

4. Système selon l'une des revendications 1 à 3 caractérisé en ce que chaque lien vocal d'un dictionnaire comporte une adresse permettant de localiser une ressource.
- 5 5. Système selon l'une des revendications 1 à 4 caractérisé en ce que chaque lien vocal d'un dictionnaire comporte un nom de la commande vocale, et en ce que le dispositif client est conçu, après réception d'un dictionnaire, pour lire et afficher les noms de tout ou partie des liens vocaux de ce dictionnaire.
- 10 6. Système selon la revendication 5 caractérisé en ce que le chaque lien vocal d'un dictionnaire comporte un attribut (« type ») permettant de spécifier si une commande vocale doit être cachée ou non, et en ce que le dispositif client (2) est conçu, après réception d'un dictionnaire, pour lire et afficher les noms uniquement des liens vocaux dont la  
15 valeur de l'attribut « type » autorise la visualisation.
7. Serveur de données comportant une unité de traitement (5) et unité mémoire (4) qui est accessible au moins en lecture par l'unité de traitement (5), et dans laquelle sont stockées une pluralité de pages destinées à être visualisées sur un dispositif client (2) après  
20 téléchargement via un réseau de télécommunication (3), caractérisé en ce que l'unité mémoire (4) comporte en lien avec chaque page au moins un dictionnaire d'un ou plusieurs liens vocaux, chaque lien vocal comportant au moins un enregistrement audio d'une commande vocale.
8. Serveur selon la revendication 7 caractérisé en ce que chaque page du  
25 serveur comporte une information identifiant le ou les dictionnaires associés.
9. Serveur selon la revendication 7 ou 8 caractérisé en ce que chaque lien vocal comporte une adresse permettant de localiser une ressource, de préférence dans l'unité mémoire (4) du serveur (1).
- 30 10. Dispositif client (2) qui d'une part est apte à échanger des données avec un serveur (1) distant et à télécharger et visualiser des pages de

- données stockées en mémoire du serveur (1), et qui d'autre part comprend des moyens (11,12) permettant l'enregistrement d'une commande vocale prononcée par un utilisateur, et des moyens de reconnaissance vocale permettant, à partir d'une commande vocale enregistrée, de déterminer et de commander automatiquement l'exécution une action associée à cette commande, caractérisé en ce que le dispositif client (2) est conçu pour télécharger en mémoire depuis le serveur (1) un dictionnaire qui est associé à une page visualisée et qui contient un ou plusieurs liens vocaux, chaque lien vocal comprenant au moins un enregistrement audio d'une commande vocale, et en ce que les moyens de reconnaissance vocale du dispositif client (2) comportent un programme de reconnaissance vocale (E) qui est conçu pour réaliser une comparaison de l'enregistrement audio correspondant à la commande vocale prononcée par un utilisateur avec le ou les enregistrements audio de chaque dictionnaire qui a été téléchargé.
11. Support mémoire (4) sur lequel est stockée au moins une page qui est visualisable sur un dispositif client (2) et qui comporte une pluralité d'instructions lisibles par le dispositif client, les instructions représentant le contenu de la page et incluant une information qui identifie au moins un dictionnaire associé à la page, ledit dictionnaire comprenant un ou plusieurs liens vocaux, un lien vocal comprenant au moins un enregistrement audio d'une commande vocale, ladite information, une fois lue par le dispositif client, déclenchant le téléchargement depuis un serveur (1) distant dudit dictionnaire.

1/8

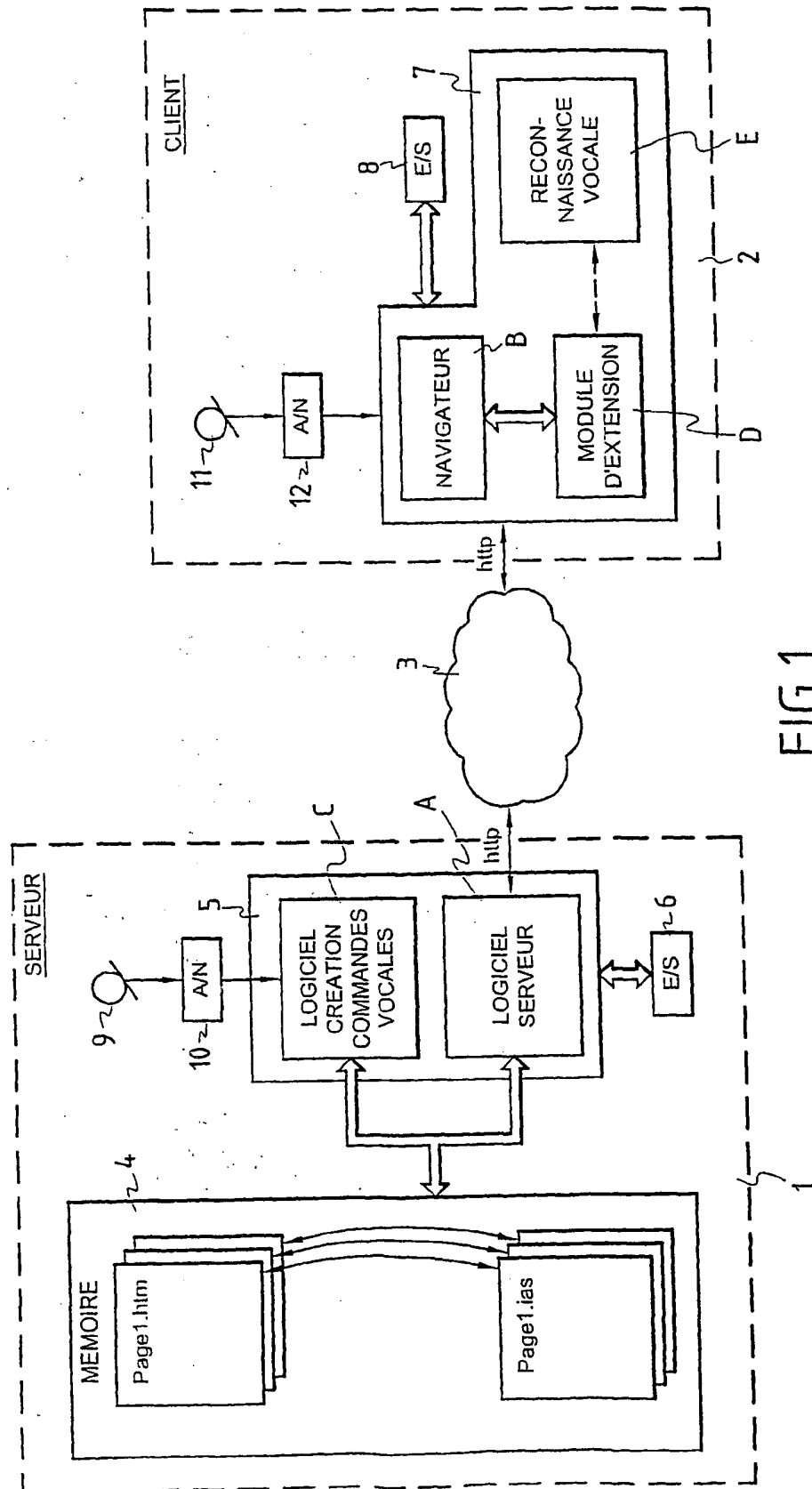


FIG.1

2/8

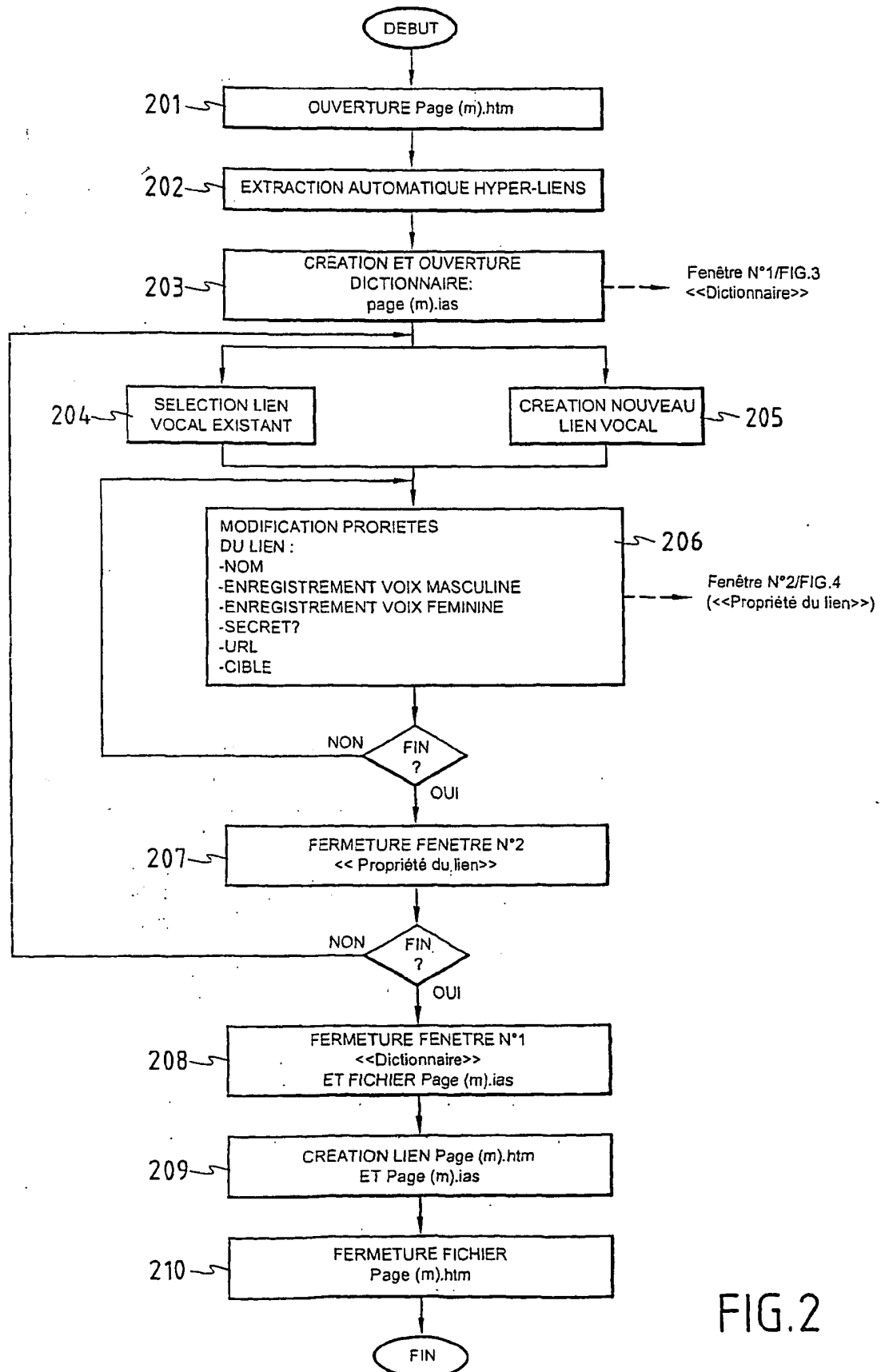


FIG.2

3/8

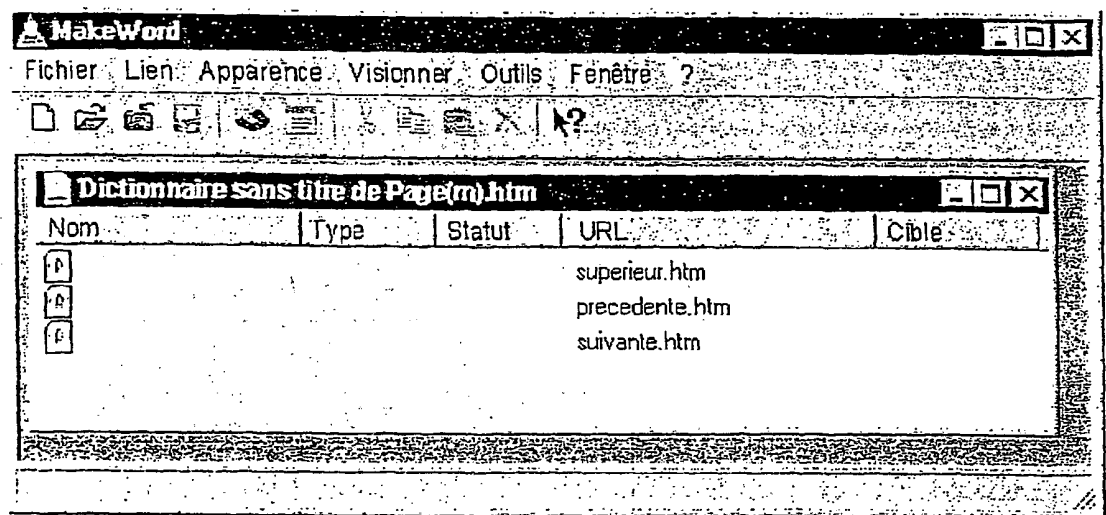


FIG.3



4/8

**Propriétés du lien** [X]

Lien vocal

Nom:

Voix masculine: Non

Voix féminine: Non

☐ Secret

Destination

URL:

Cible:

FIG.4

5/8

**Propriétés du lien** [X]

**Lien vocal**

Nom:

Voix masculine: oui

Voix féminine: Oui

☐ Secret

**Destination**

URL:

Cible:

FIG.5

6/8

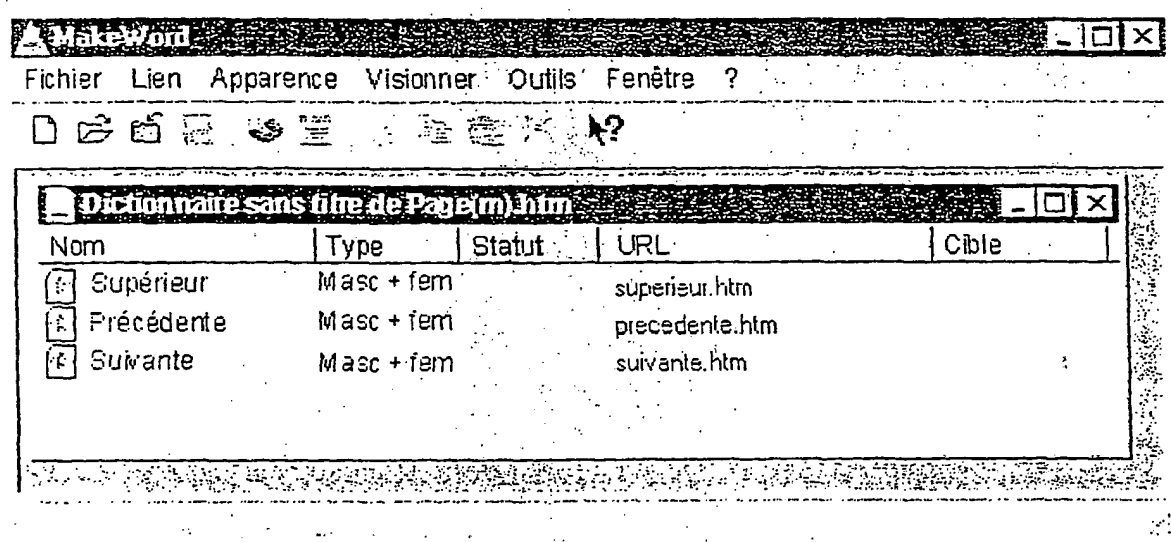


FIG.6

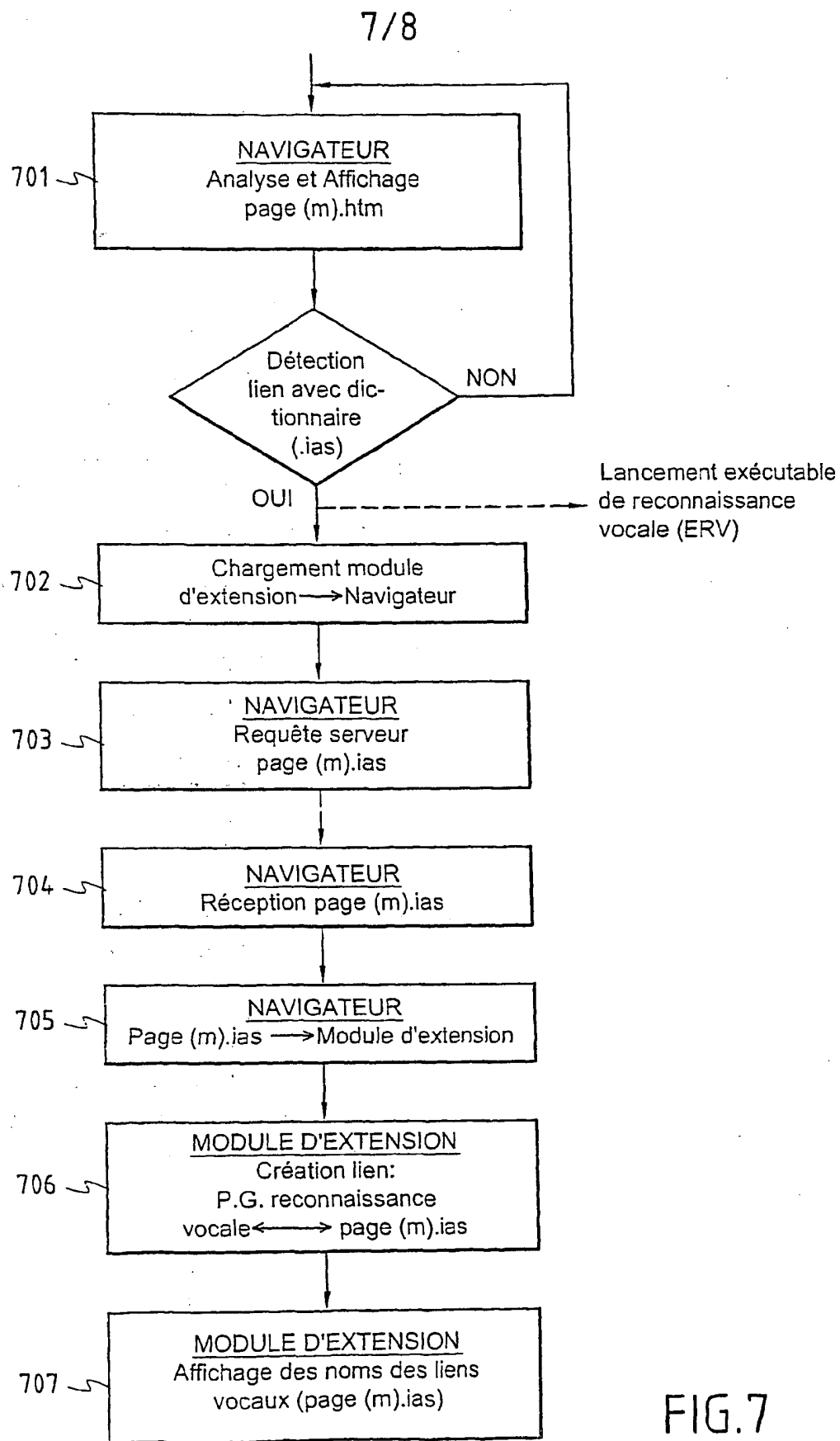


FIG. 7

8/8

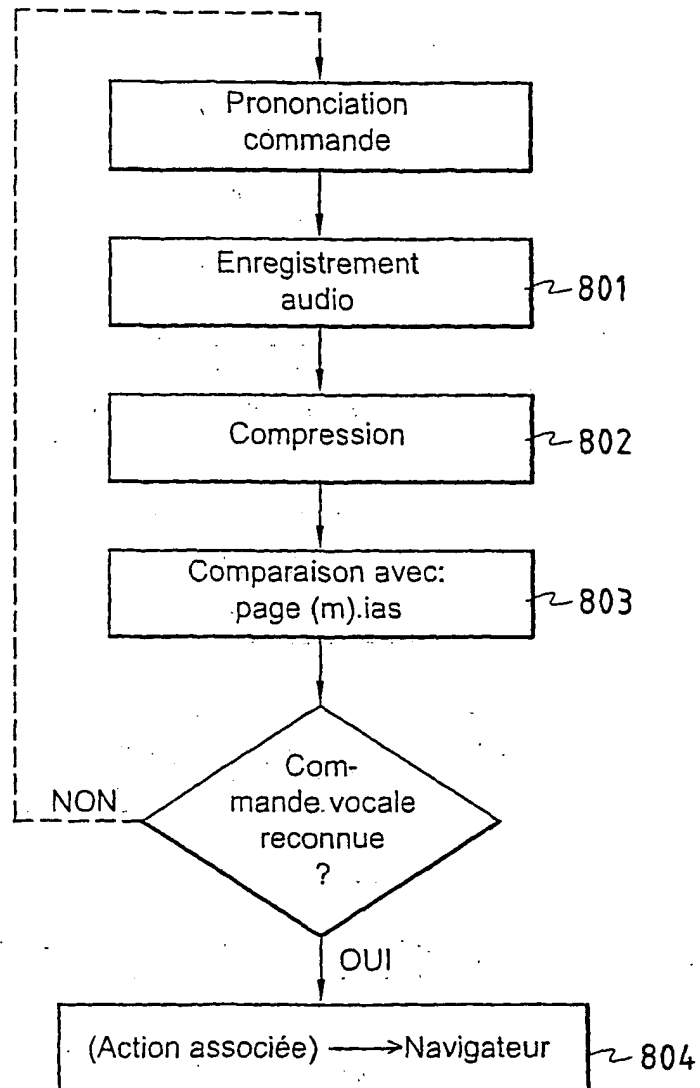


FIG.8

## INTERNATIONAL SEARCH REPORT

International Application No  
PCT/FR 01/01560

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 7 G06F3/16

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G06F G10L H04M

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category * | Citation of document, with indication, where appropriate, of the relevant passages   | Relevant to claim No. |
|------------|--|-----------------------|
| A          | US 5 355 433 A (YASUDA SEIGOU ET AL)<br>11 October 1994 (1994-10-11)<br>cited in the application<br>the whole document   | 1-11                  |
| A          | EP 0 847 179 A (AT & T CORP)<br>10 June 1998 (1998-06-10)<br>page 2, line 3 - line 44; figures 1,2   | 1,7,10,<br>11         |
| A          | WO 99 48088 A (BROWN N GREGG ;INROAD INC<br>(US); PROFIT JACK H JR (US))<br>23 September 1999 (1999-09-23)<br>page 6, line 29 -page 9, line 22; figures<br>2,3 | 1,7,10,<br>11         |

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

\* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*G\* document member of the same patent family

Date of the actual completion of the international search

31 August 2001

Date of mailing of the international search report

07/09/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Moens, R

## RAPPORT DE RECHERCHE INTERNATIONALE

Der » Internationale No  
PCT/FR 01/01560A. CLASSEMENT DE L'OBJET DE LA DEMANDE  
CIB 7 G06F3/16

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

## B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)  
CIB 7 G06F G10L H04M

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si réalisable, termes de recherche utilisés)

EPO-Internal

## C. DOCUMENTS CONSIDERES COMME PERTINENTS

| Catégorie * | Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents   | no. des revendications visées |
|-------------|--|-------------------------------|
| A           | US 5 355 433 A (YASUDA SEIGOU ET AL)<br>11 octobre 1994 (1994-10-11)<br>cité dans la demande<br>le document en entier  | 1-11                          |
| A           | EP 0 847 179 A (AT & T CORP)<br>10 juin 1998 (1998-06-10)<br>page 2, ligne 3 - ligne 44; figures 1,2   | 1,7,10,<br>11                 |
| A           | WO 99 48088 A (BROWN N GREGG ; INROAD INC.<br>(US); PROFIT JACK H JR (US))<br>23 septembre 1999 (1999-09-23)<br>page 6, ligne 29 -page 9, ligne 22;<br>figures 2,3 | 1,7,10,<br>11                 |

☐ Voir la suite du cadre C pour la fin de la liste des documents☒ Les documents de familles de brevets sont indiqués en annexe

## \* Catégories spéciales de documents cités:

- \*A\* document définissant l'état général de la technique, non considéré comme particulièrement pertinent
- \*E\* document antérieur, mais publié à la date de dépôt international ou après cette date
- \*L\* document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)
- \*O\* document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens
- \*P\* document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

- \*T\* document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention
- \*X\* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément
- \*Y\* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier
- \*Z\* document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

31 août 2001

Date d'expédition du présent rapport de recherche internationale

07/09/2001

Nom et adresse postale de l'administration chargée de la recherche internationale  
Office Européen des Brevets, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Fonctionnaire autorisé

Moens, R

# INTERNATIONAL SEARCH REPORT

Information on patent family members

Intel International Application No

PCT/FR 01/01560

| Patent document<br>cited in search report | Publication<br>date | Patent family<br>member(s)                   | Publication<br>date                    |
|---|---------------------|--|--|
| US 5355433 A                              | 11-10-1994          | JP 2989211 B<br>JP 3274594 A<br>DE 4109785 A | 13-12-1999<br>05-12-1991<br>02-10-1991 |
| EP 0847179 A                              | 10-06-1998          | CA 2213591 A<br>JP 10207685 A                | 04-06-1998<br>07-08-1998               |
| WO 9948088 A                              | 23-09-1999          | AU 3104599 A                                 | 11-10-1999                             |

Form PCT/ISA/210 (patent family annex) (July 1992)



# RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Der. e Internationale No

PCT/FR 01/01560

| Document brevet cité<br>au rapport de recherche | Date de<br>publication | Membre(s) de la<br>famille de brevet(s)      | Date de<br>publication                 |
|---|------------------------|--|--|
| US 5355433 A                                    | 11-10-1994             | JP 2989211 B<br>JP 3274594 A<br>DE 4109785 A | 13-12-1999<br>05-12-1991<br>02-10-1991 |
| EP 0847179 A                                    | 10-06-1998             | CA 2213591 A<br>JP 10207685 A                | 04-06-1998<br>07-08-1998               |
| WO 9948088 A                                    | 23-09-1999             | AU 3104599 A                                 | 11-10-1999                             |

Formulaire PCT/ISA/210 (annexe familles de brevets) (juillet 1992)

THIS PAGE BLANK (USPTO)